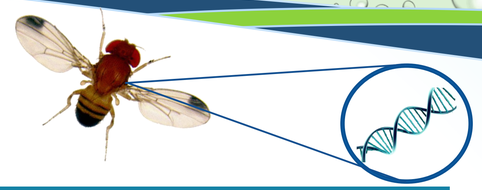
In Pursuit of Phase Separation

Understanding the amino acid compositional bias of intrinsically disordered proteins in *D. melanogaster*

Prepared by Elvira Mukharryamova; Supervised by Dr. Don Nelson and Dr. Eric Bottos
Thompson Rivers University, Department of Biological Sciences

Abstract

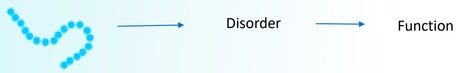
The ability of linear chains of amino acids to fold into the 3-dimensional structure governs biological function adhering to the structure-function postulate. The proteins that lack a three-dimensional shape or contain regions of intrinsic disorder within its sequence are called intrinsically disordered proteins (IDPs). IDPs are associated with large fraction of charged, hydrophilic and disorder-promoting amino acids, which deprive the protein of the ability to fold without impairing its functional state. Many proteins were shown to undergo phase separation in vitro and in vivo over the years; however, the molecular processes that govern phase transitions in eukaryotic cells remain largely unknown. In this project, the bioinformatic analysis of intrinsically disordered proteins in *Drosophila melanogaster* was performed to examine distribution of the disordered content and the amino acid compositional bias. Genetic line diagrams and a primer design for the selected proteins were completed as part of the future work analysis. The findings of the bioinformatic analysis and potential laboratory experiment will contribute to our understanding of molecular principles behind phase separation and its application in human disease.



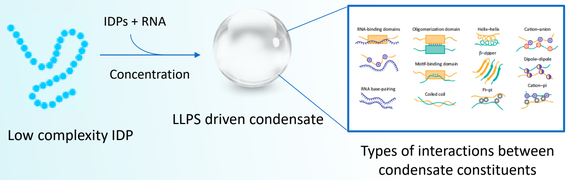
Introduction

In cells, proteins carry out a myriad of vital functions including movement, catalysis, storage, signaling, and others. This vast functional repertoire is directly dependent on protein's 3-dimensional structure, which, in turn, is dictated by its amino acid composition.

At the end of the 20th century, a new class of novel proteins, called **intrinsically disordered proteins (IDR)**[5] was discovered. IDPs' are unable to fold into a 3-D conformation, but fully remain their functional capabilities.



IDPs interact with each other and different cytoplasmic biomolecules via weak multivalent interactions. Once the concentration of IDPs at one location reaches a concentration threshold, IDPs form dynamic liquid droplets through the process of **liquid-liquid phase separation (LLPS)**[4]. LLPS driven droplets enable selective compartmentalization of molecules that can impede chemical reactions inside eukaryotic cells.



Why is it Relevant?

IDPs and IDR containing proteins were shown to participate in regulatory processes such as post-translational modifications, transport of molecules, signal transduction and regulation of gene expression[4]

The aberrant phase separation was linked to the formation of protein plaques and tangles that are associated with the following diseases:

- Alzheimer's
- Spinocerebellar Ataxia
- Cancer[4].

If we understand the molecular mechanisms of LLPS driven condensates, we could potentially mitigate disease.

Objective: To characterize and classify intrinsically disordered proteins in *Drosophila melanogaster* according to the percent of disorder and amino acid composition. Also, to identify the distribution of IDRs in proteins for future laboratory analysis.

Methodology and Results

Part I. *Drosophila m.* proteome screening

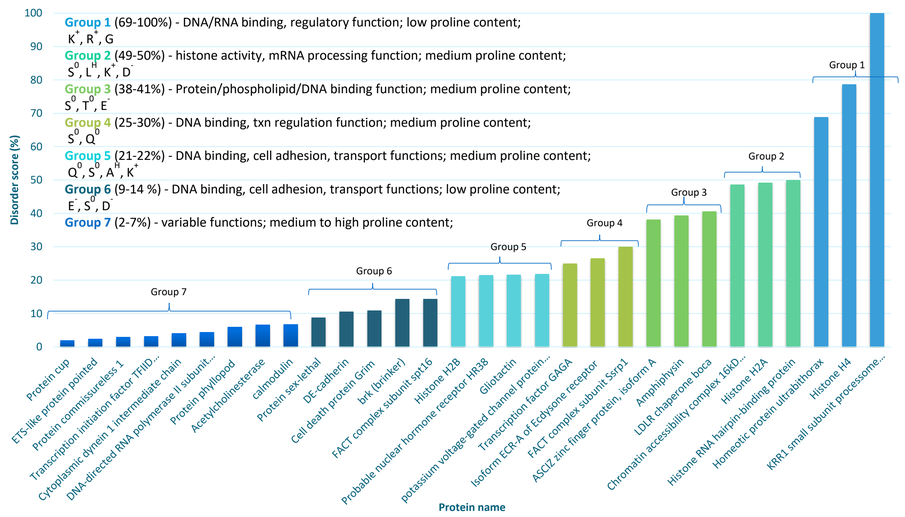
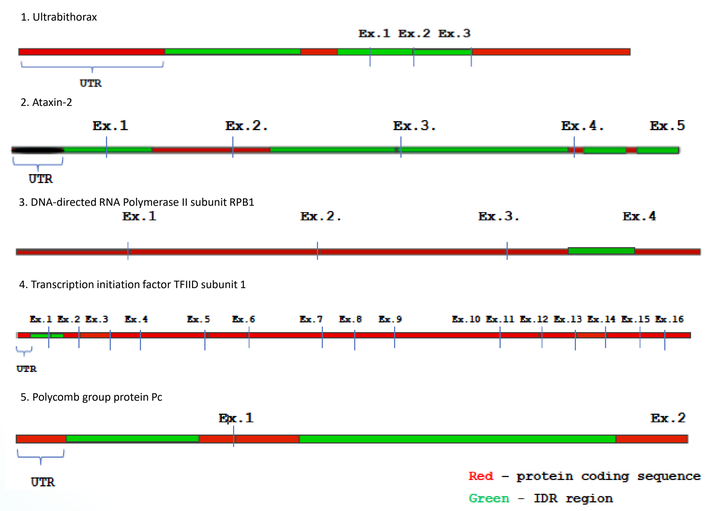


Figure 1: The degree of disorder as a percentage of intrinsically disordered residues for curated proteins of Drosophila Melanogaster obtained from the DisProt database.

Key observations:

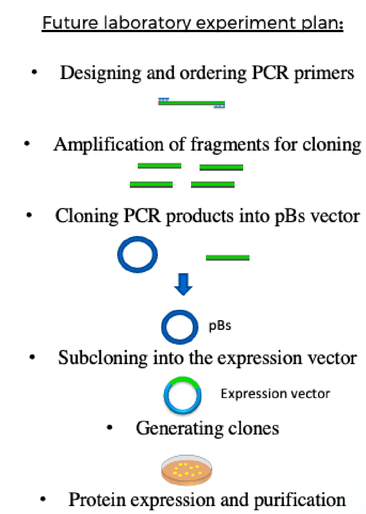
- 1. IDPs with the highest percent disorder are associated with ribosome biogenesis, nucleosome activity and regulation of gene expression
- 2. As the percent of disorder decreases, IDPs demonstrate higher functional variability.
- 3. The common characteristics of *Drosophila's* 54 disordered proteins include low-to-medium sequence complexity and low hydrophobicity.
- Proline, which is a disorder-promoting amino acid, appears with variable frequencies. However, the residue was observed to be more abundant in IDPs with low disorder content.
- Polar and charged amino acids (R, K, S, Q) as well as the disorder-promoting amino acids (P, G) were in high abundance.
- In contrast, non-polar amino acids (C, I, M, W, Y, V) were shown to be least abundant in IDPs.

Part II. IDR distribution in mRNA sequences of 5 selected IDPs



Future laboratory experiment plan:

- Designing and ordering PCR primers
- Amplification of fragments for cloning
- Cloning PCR products into pBs vector
- Subcloning into the expression vector
- Generating clones
- Protein expression and purification



Primer Design

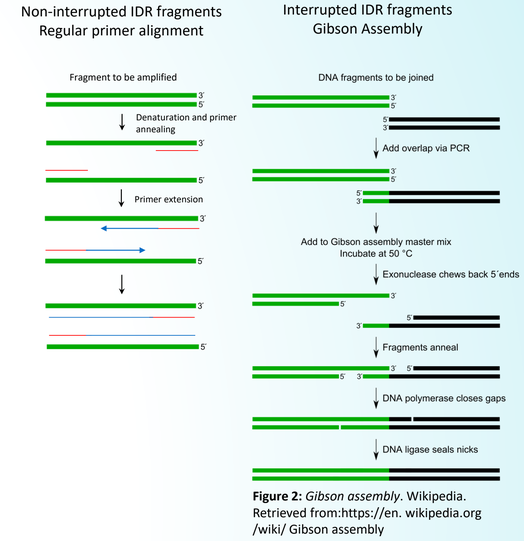


Figure 2: *Gibson assembly*. Wikipedia. Retrieved from:https://en. wikipedia.org /wiki/ Gibson assembly

Conclusion

- The amino acid composition of IDPs in *Drosophila melanogaster* demonstrated compositional bias that contributes to the inability of IDPs to form a hydrophobic core but maintain linear structure.
- The location of IDRs of five selected proteins was visualized on the line graph as part of the future work analysis. According to the results, all 5 proteins could be used for *in vitro* phase separation laboratory experiment since they all contain long IDRs and were experimentally shown to undergo LLPS.
- Although, the research concerning IDPs is undergoing a true progress, there are more questions than answers regarding the intracellular functions and the modes of condensate regulation. The association of IDP proteins with several incurable human diseases like cancer, diabetes and several neurodegenerative disorders has already been established; however, the role of unstructured proteins in the disease progression remains unexplored.
- Knowing that the solid protein aggregates are correlated with disease, being able to reverse the solid state back into its dynamic form provides new therapeutic targets for disease treatments.